Carlos Enmanuel Soto Lopez
Via Beirut 2
Trieste, Italy
I-34014
csoto@ogs.it

Dear Reviewer,

We appreciate your comments on the manuscript, please find attached a point-by-point response to your comments, together with a description of the changes made to the original manuscript.

I apologize for not attaching the revised version here, as there is a note on the discussion page requesting that updated versions of the paper not be posted in that section. I will send the revised manuscript to the editorial office so they can upload it in the appropriate location.

Yours sincerely,

Carlos Enmanuel Soto Lopez

Point by point reply

Reviewer comment: *Line 89, "data filtered with an analog high pass filter, using the package SciPy". It cannot be an analog filter. Maybe the prototype is analog, but it must be a digital filter. Please explain/correct this*

Author reply: We substituted the word 'analog' with the specific name of the filter. Now the text say "...were filtered with a Butterworth high-pass filter, using the package SciPy".

Reviewer comment: *Line 130, respectably =¿ respectively*

Author reply: I fixed the typo. Thank you for pointing it out.

Reviewer comment: *Line 249, in eq. 19 K is not defined.*

Author reply: K is the Jacobian of R with respect to z. We added the information in the new version of the manuscript.

Reviewer comment:
- *Line 313, DK divergence => DKL divergence.*
- *Line 347, the DK divergence => the DKL divergence.*
- *Line 357, Cholezky => Cholesky.*
- *Line 371, q(z,y,x) => q(z—y,x)*

Author reply: We corrected all the typos in the updated manuscript.

Reviewer comment: *Fig. 1, What is exactly "model of the measurements". Maybe you can add* R_{rs}^{MODEL} *and* $\hat{H}(Z, X; \Lambda)$.

Author reply: We changed the term to "Forward Model", as this is the name used throughout the rest of the paper.

Reviewer comment: *Fig. 2 caption:* $q_\phi(z|z, y)$ *should be* $q_\phi(z|x, y)$*, and "dose that learn the Cholezky" =¿ those? that learn the Cholesky.*

Author reply: We changed the typos in the updated manuscript.

Reviewer comment: *Fig 5. (Related to it) Please give more explanation on: Why MCMC sub or over estimates? Why the SGVB does not have a confidence interval if the method provides it? And why do you think is the cause the SGVB uncertainty fails?*

Author reply: Regarding the confidence interval, the Variational Bayes method approximates the posterior of the optical constituents, called Z, or latent variables in the text.To this end, the neural network estimates both the mean and the covariance; however, since the covariance underestimates the true discrepancy between predictions and observations, we relied only on the mean values. For uncertainty estimation, linearizing the Forward Function around the estimated mean and applying standard error propagation methods yielded better performance. On the other had, the Variational Bayes estimates only the MLE of the parameters, as plotted in Fig. 5.

Regarding the different values obtained by the two methods: after standardizing the observations (dividing by the standard deviation) used for the MCMC algorithm, the differences are minimal (see Fig. 10 of the new manuscript). They could be due, e.g, to overfitting, changes in the loss function (the NN loss function used a regularization), or model error, i.e. the error generated by the approximation the posterior with a NN.

Regarding the underestimation of the covariance matrix, I speculate that it is due to training the neural network to match observations directly. The original goal is to invert the Forward Model.

If we were to use the Maximum Likelihood Estimate (MLE), computing the inverse of the Fisher Information Matrix would result in extremely large uncertainty estimates. In contrast, the Maximum A Posteriori (MAP) estimate yields lower uncertainty, as it incorporates prior information. The 'posterior' represented by the neural network includes not only a term for the observations but also a regularization term, which—when viewed from a Bayesian perspective—can be interpreted as an additional form of prior information. This added prior information reduces the spread of the posterior, thereby underestimating the true discrepancy between the modeled data and the observations. However, since this remains speculative, in the paper we focused solely on reporting the result that the neural network underestimates the uncertainty.